DOCUMENT RESUME

ED 295 073 CG 020 813

AUTHOR Elmore, Patricia B.; And Others
TITLE Using Tests Responsibly in Counseling.
PUB DATE Mar 88
NOTE 20p.; Paper presented at the Annual Convention of the American Association for Counseling and Development (Chicago, IL, March 20-23, 1988).
PUB TYPE Speeches/Conference Papers (150) -- Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Counseling Services; *Counselor Role; *Educational Testing; Elementary Secondary Education; Higher Education; *Psychological Testing; School Counseling; School Counselors; Social Discrimination; *Test Bias; *Test Validity
IDENTIFIERS *Fairness

ABSTRACT

This document provides a discussion of the scrutiny for bias and unfairness that standardized tests have come under by persons involved in the construction, administration, use, and interpretation of test results. After this discussion six simulation exercises are provided. In these exercises a situation is described involving the use of test scores by a counselor and then questions are asked about possible test bias, fairness, score differences, and counselor action. Eleven standards for educational and psychological testing and bias in measurement taken from the Standards for Test Use in Counseling are given. Five guidelines for responsible test use in counseling discussed include: (1) using the test for the purpose it was intended; (2) avoiding using a single test score to make a decision about an individual; (3) using caution when tests have different norms for different groups of individuals; (4) using caution when a procedure assigns individuals to groups on the basis of test score; and (5) being prepared to investigate test score differences across groups of students. Counselors are advised to use common sense above all. (ABL)

Using Tests Responsibly in Counseling

Patricia B. Elmore
Southern Illinois University

Esther E. Diamond
Educational and Psychological Consultant
Evanston, Illinois

Ruth B. Ekstrom
Educational Testing Service

Paper presented at the Annual Meeting of the
American Association for Counseling and Development
Chicago

March, 1988

Using Tests Responsibly in Counseling[1]

Standardized tests and their uses have come under increasing scrutiny by all concerned with the construction, administration, use, and interpretation of test results. Tests have become the target of federal and state legislation designed to ensure fairness to women and men, to ethnic and minority groups, and to handicapped persons (Berk, 1982).

The importance of fairness in testing cannot be overestimated; equity in education and employment is not possible unless there is equity in the measures used to make educational and vocational decisions about individuals and groups. Progress from grade to grade in school, admission to particular school programs and to institutions of higher education, counseling students and adults about educational and vocational plans, and selection for employment, job placement, and promotion are all affected by tests of various kinds (Diamond & Tittle, 1985). The instruments used for these purposes are achievement tests, aptitude tests, and interest inventories.

Bias in testing became a concern of the Association for Measurement and Evaluation in Guidance (AMEG) (now the Association for Measurement and Evaluation in Counseling and Development [AMECD]) in the early 1970s, when the issue of sex bias in interest measurement was raised by several members. The AMEG Commission on Sex Bias in Measurement was formed to study the matter and published a report in 1973 (AMEG Commission on Sex Bias in Measurement, 1973). Subsequent changes that

publishers made or planned to make indicated that the study had had sizable impact (AMEG Commission on Sex Bias in Measurement, 1977).

In 1977, the AMEG Commission on Sex Bias in Interest Measurement expanded its charge to include an investigation of what publishers of widely used standardized achievement tests were doing to identify and minimize sex bias in these tests. The findings (Diamond, 1980), as reflected in test manuals and reports, indicated that publishers paid greater attention to procedures for minimizing test bias in tests published after 1970 than to those published before that time.

In 1983 the AMECD Committee on Bias in Measurement surveyed publishers of standardized achievement tests to determine the techniques currently used to minimize not only sex bias but also ethnic and minority bias in standardized achievement tests. The survey results were published in 1986 (Diamond & Elmore, 1986). It was concluded that the effort to detect and minimize test bias must be a dynamic, on-going process and the joint responsibility of all stakeholders, including those who develop the tests, those who publish them, those who select them and use the results, and those who are affected by the way in which the results are used.

This presentation will include a discussion of responsible use and interpretation of achievement, aptitude, and interest tests in a counseling setting. Particular attention will be given to discussion of counselors' opinions of how inappropriate use of tests is related to test bias. Simulation exercises will be presented in which tests could

be misused or test results interpreted inappropriately. The audience will participate by answering questions for each simulation. To summarize, the presenters and audience will jointly develop the steps a counselor should follow in considering tests which show score differences for majority and minority groups of students.

Definitions

Through the years consensus seems to have been reached on the definition of test bias and the distinction between bias and unfairness in testing (Diamond & Tittle, 1985).

Bias: refers to the intrinsic characteristics of a test; that is, to "content, the construct or constructs the test is supposedly measuring, and the context within which the content is embedded" (p. 168). Bias is considered to occur "when two individuals of equal ability but from different groups respond differently to a test item and therefore do not have the same probability of success on the item" (p. 168).

Unfairness: refers to "ethical questions involving use of the test results" (p. 168).

Simulation Exercises[2]

Exercise 1

An elementary school counselor notices that on an achievement test there is a large difference between the scores of Black and White students on the Word Problems subtest of the Mathematics portion of the test. The counselor gives these same test problems orally to a group of students and finds no difference between the proportion of White and Black students solving the problems correctly.

1. Is the test biased or unbiased?

   Biased Don't Know Unbiased

2. Is the test fair or unfair?

   Fair Don't Know Unfair

3. What might explain the score differences?

   _____

   _____

   _____

   _____

4. What should the counselor do?

   _____

   _____

   _____

   _____

Exercise 2

A high school counselor is asked by the principal to give a verbal analogies test to two groups of students to determine their reasoning ability. Both groups of students came to the United States three years ago and now read English about equally well. One group is from Japan and one is from Italy. The counselor reviews the content of the test and notes that 80 percent of the items are based on words with Latin origins.

1. Is the test biased or unbiased?

   Biased Don't Know Unbiased

2. Is the test fair or unfair?

   Fair Don't Know Unfair

3. What might explain the score differences?

   _____

   _____

   _____

   _____

4. What should the counselor do?

   _____

   _____

   _____

   _____

Exercise 3

A newly appointed college counselor notices that twice as many males as females receive a special state scholarship for outstanding students. When she inquires, she finds that the students are selected by adding the Verbal and Mathematics scores on the college's admission test, and then taking the students with the highest combined scores.

1. Is the test biased or unbiased?

   Biased Don't Know Unbiased

2. Is the test fair or unfair?

   Fair Don't Know Unfair

3. What might explain the score differences?

   _____

   _____

   _____

   _____

4. What should the counselor do?

   _____

   _____

   _____

   _____

Exercise 4

A junior high counselor is confronted by a distraught seventh grade student. The student was selected to participate in the statewide Talent Search during sixth grade because she had regularly scored at the 95th percentile on standardized achievement tests during elementary school. As part of the Talent Search the student was administered a college admission test usually given to high school juniors and seniors. The student scored below the 25th percentile compared to other Talent Search students on the college admission test and feels she is a failure.

1. Is the test biased or unbiased?

Biased Don't Know Unbiased

2. Is the test fair or unfair?

Fair Don't Know Unfair

3. What might explain the score differences?

_____

_____

_____

_____

4. What should the counselor do?

_____

_____

_____

_____

Exercise 5

Fran, a college freshman, has taken an interest inventory that compares her responses with those of both men and women in a variety of different occupations. However, some of the occupations have been normed only on men and some only on women. In general, Fran's highest scores are those on women's norms, for occupations such as teacher, librarian, and social worker--occupations in which women have generally predominated. Her highest scores on male norms, however, are for occupations such as personnel manager, industrial psychologist, and architect, for which female norms have not yet been developed.

1. Is the inventory biased or unbiased?

   Biased Don't Know Unbiased

2. Is the inventory fair or unfair?

   Fair Don't Know Unfair

3. Why do you think Fran's scores on male norms are lower than those she obtained on female norms?

4. How can the counselor help Fran use her scores in educational and vocational planning?

Exercise 6

Lisa has just taken an interest inventory that measures vocational interests in a number of broad areas. Lisa's responses are compared with those of girls at her grade level (ninth). Her score on the Scientific scale is at the 71th percentile. On the Mechanical scale it is at the 73rd percentile. Her highest score is on the Clerical scale. The counselor comments that boys with the same raw scores on the Scientific and Mechanical scales would score at much lower percentiles, but on the Clerical scale their scores would be at a higher percentile.

1. Is the inventory biased or unbiased?

   Biased Don't Know Unbiased

2. Is the inventory fair or unfair?

   Fair Don't Know Unfair

3. What do you think might explain the differences between Lisa's scores on the two sets of norms (compared with girls and compared with boys)?

   _____

   _____

   _____

4. What additional information about the scores might help Lisa?

   _____

   _____

   _____

5. How can the counselor help Lisa use her scores in educational and vocational planning?

   _____

   _____

   _____

Standards for Educational and Psychological Testing [3]

Standards for Test Use in Counseling (pp. 56-58)

Standard 9.1 Testing for counseling should have as its primary goals the acquisition of relevant information and the reporting of that information with apppropriate interpretations so that clients from diverse backgrounds can be assisted in making important educational, personal, and career decisions. (Primary)

Standard 9.2 Counselors should review the interpretive materials provided to clients to evaluate accuracy, clarity, and usefulness of the materials. Manuals for tests or computer-based interpretations should be evaluated for evidence for the validity of specific interpretations made. (Primary)

Standard 9.3 Counselors should review technical data and develop a rationale for the decision to use combined or separate norms for females and males in reports to test takers. (Primary)

Standard 9.4 If a publisher packages tests that are to be used in combination for counseling, the counselor should review the manual for the rationale for the specific combination of tests used and the justification of the interpretive relationships among the scores. (Primary)

Standard 9.5 Counselors should examine test manuals for any available information about how suggested or implied career options (i.e., the vocational program or occupation suggested by the highest scores on the test) are distributed for samples of the typical respondents of each gender and relevant racial or ethnic groups. (Primary)

Standard 9.6 Counselors should review the test materials that are provided to the test takers to be sure that such materials properly caution the test taker not to rely on the test scores solely when making life-planning decisions. The counselor should encourage the test taker to consider other relevant information on personal and social skills, values, interests, accomplishments, experiences, and on other test scores and observations. (Primary)

Standard 9.7 Counselors should encourage multiple valid assessments of an individual's abilities, social skills, and interests. (Primary)

Standard 9.8 Counselors should review the interpretive materials for ability or interest measures and for other tests that are used with people who are reentering employment or education or changing work settings for their appropriateness for these clients. A counselor should consider the age, experience, and background of the client as they are compared with the characteristics of the norm groups on which the scores are based. (Primary)

Standard 9.9 Counselors should review interpretive materials for tests to ensure that case studies and examples are not limited to illustrations of people in traditional roles. (Secondary)

Standards for Bias in Measurement

Standard 1.20 Investigations of criterion-related validity for tests used in selection decisions should include, where feasible, a study of the magnitude of predictive bias due to differential prediction for those groups for which previous research has established a substantial prior probabilty of differential prediction for the particular kind of test in question. (Conditional) (p.17)

Standard 3.10 When previous research indicates the need for studies of item or test performance differences for a particular kind of test for members of age, ethnic, cultural, and gender groups in the population of test takers, such studies should be conducted as soon as is feasible. Such research should be designed to detect and eliminate aspects of test design, content, or format that might bias test scores for particular groups. (Conditional) (p. 27)

Guidelines for Responsible Test Use in Counseling

1. Be sure you are using a test for the purpose(s) for which it is intended. Check in the test manual to see what the validity evidence is. What validity criterion was used? Is this the same outcome you are trying to determine? Is there validity evidence for the same kinds of individuals you will be testing?

   Any test you use should provide evidence of the intended use(s) of the test scores. Be sure that the validity evidence reported has been collected in the same situation (or a very similar one) as the situation in which you will use the test. Be sure there is validity evidence for individuals like those you will be testing. Ideally, the test manual should tell you more about the group that was tested than race/ethnicity and sex. For example, what school grade, school curriculum, or job were they in?

2. Avoid using a single test score to make a decision about an individual.

   Supplement test scores with other relevant information about individuals. For example, in college admissions decisions, school grades should be given at least as much weight as test scores. Whenever possible, use several test scores rather than one score. Ideally individuals should have the opportunity to be retested whenever they feel that a test score is not an accurate indicator of their abilities or interests. Alternatively, individuals might be given the opportunity to demonstrate the same skill or level of interest in a different, non-test situation.

3. Be cautious about using tests which have different norms for different groups of individuals (e.g., separate norms for females and males). Be sure you understand the test maker's rationale for providing different norms. Be sure that you do not make erroneous assumptions based on separate norms.

   For example, females tend to take less mathematics in high school than males. Minority students are more likely to be enrolled in non-academic high school curricula than are white students. Separate norms can help counselors take account of such differences. Ideally, it would be better if there were norms based on the actual differences (number of mathematics courses taken, high school curriculum) than to use sex or race norms, which are based on the fact that sex and race are correlated with these real differences.

4. Be cautious using any procedure which assigns individuals to groups on the basis of a test score. Review the procedure at regular intervals to be sure you are not operating on invalid or outdated assumptions.

   For example, some state scholarship programs have operated for years on the assumption that females score higher on the verbal portion of college entrance tests and males score higher on the mathematics portion, thus producing similar total scores. This has not been true for more than 20 years.

   Check to see if the assignment criterion is still appropriate.

   If possible, collect evidence of individual performance to support the assignments or to provide information for reconsideration of the current procedure.

   Check classification rates by subgroups (race/ethnicity, sex, etc.). If there are differences, try to determine if these are valid. For example, test scores are sometimes used to assign high school students to a curriculum track; minority students tend to be assigned to non-academic curricula more often than white students. In this case it might be appropriate to find out if using teacher recommendations and/or junior high school grades instead of or in addition to the test scores would help provide more equitable curriculum assignments.

   Be especially cautious when using pass-fail test scores or cut scores. Ideally, these should be used only when there is no other alternative. The use of such scores should be supported by statistical data showing a clear relationship between the test scores and the outcome. Individuals should have opportunities for retesting when pass-fail or cut scores are used.

5. Be on the lookout for test score differences across groups of students. Be prepared to investigate the reasons for these differences. It is especially important to differentiate between problems that are due to the content of the test that is being used and problems that are due to different learning opportunities for various groups of students. An outline of steps a counselor might take in reviewing tests which show score differences in results for different groups of students, to ensure that the test is fair and will be used responsibly, follows:

First, inspect the content of the test. What kinds of subjects are being covered?

Second, look at enrollment data for the courses most closely related to the test content. Are equal numbers of students from each group taking and passing these courses? If so, the test may be biased. Contact the test publisher or read the test manual to see what evidence is available about the use of the test with different groups of students.

Third, if there are different enrollment patterns for different groups of students, try to determine the reasons for these course-taking differences. Then think of some interventions that may reduce these differences. For example, if the test score differences seem to be related to different rates of enrollment in mathematics and science courses, you may want to use some of the excellent counseling materials that are available to encourage minorities and women to take more science and mathematics. Do you and other counselors work actively to encourage students to take a strong educational program?

Fourth, when using test results for decision making, be sure to take into consideration the factors which influence these results, such as different opportunities to learn. A low score on an algebra test by a student who has not studied algebra does not mean that this student is unable to learn algebra.

Above all, use your common sense. As counselors, you have good insights into people, their abilities, interests and motivations. Don't use any test score that you feel is unfair to your client.

References

AMEG Commission on Sex Bias in Measurement. (1973). AMEG Commission report on sex bias in interest measurement. Measurement and Evaluation in Guidance, 6, 171-177.

AMEG Commission on Sex Bias in Measurement. (1977). A case history of change: A review of response to the challenge of sex bias in career interest inventories. Measurement and Evaluation in Guidance, 10, 148-152.

American Educational Research Association/American Psychological Association/National Council on Measurement in Education. (1985). Standards for educational and psychological testing. Washington, DC: American Psychological Association, Inc.

American Personnel and Guidance Association. (1980). Responsibilities of users of standardized tests. Falls Church, VA: Author.

Berk, R. A. (Ed.). (1982). Handbook of methods for detecting test bias. Baltimore: The Johns Hopkins University Press.

Diamond, E. E. (1980). The AMEG Commission Report on sex bias in achievement testing. Measurement and Evaluation in Guidance, 13, 135-147.

Diamond, E. E., & Elmore, P. B. (1986). Bias in achievement testing: Follow-up report of the AMECD Commission on Bias in Measurement. Measurement and Evaluation in Counseling and Development, 19, 102-112.

Diamond, E. E., & Tittle, C. K. (1985). Sex equity in testing. In S. Klein (ed.), Handbook for achieving sex equity through education. Baltimore: Johns Hopkins University Press.

Footnotes

[1]Portions of pages 2 and 3 of this manuscript are adapted from "Bias in achievement testing: Follow-up report of the AMECD Commission on Bias in Measurement" by E. E. Diamond and P. B. Elmore, 1986, Measurement and Evaluation in Counseling and Development, 19, p. 102-103. Copyright 1986 by the American Association for Counseling and Development. Adapted by permission of the American Association for Counseling and Development.

[2]Exercises 1 and 2 on pages 6 and 7 of this manuscript are adapted from "Definitions of bias" by L. A. Shepard, 1982, in R. A. Berk (Ed.), Handbook of Methods for Detecting Test Bias (pp. 11-12), Baltimore, MD: The Johns Hopkins University Press. Copyright 1982 by The Johns Hopkins University Press.

[3]Pages 12 and 13 of this manuscript are from Standards for Educational and Psychological Testing (pp. 56-58, 17, 27) by American Educational Research Association/American Psychological Association/National Council on Measurement in Education, 1985, Washington, DC: American Psychological Association, Inc. Copyright 1985 by the American Psychological Association. Reproduced by permission of the American Psychological Association.